PATENT ABSTRACTS OF JAPAN

(11)Publication number:

09-146842

(43)Date of publication of application: 06.06.1997

(51)Int.CI.

G06F 12/08

G06F 12/08

G06F 3/06

G06F 3/06

(21)Application number: 07-300967

(71)Applicant: HITACHI LTD

(22)Date of filing:

20.11.1995

(72)Inventor: KOBAYASHI RIE

MATSUMOTO YOSHIKO

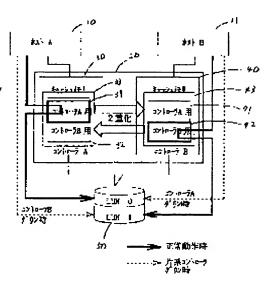
MURAOKA KENJI

(54) STORAGE SUBSYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide the storage subsystem with a controller which eliminates exclusive control over a cache between plural controllers sharing the cache.

SOLUTION: The cache areas of caches 33 and 43 in which data are written mutually in multiple are divided by processors and the controllers 30 and 40 access only their controller control areas. The cache areas that the controllers use are fixed to eliminate the need for exclusive control between the processors and prevent deterioration in performance due to multiprocessor constitution.



LEGAL STATUS

[Date of request for examination]

31.10.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本四十十 () P)

(12)公開特許公報(A)

(11)特許出版公司等号 特開平9-146842

★間平9-146842

(43)公額日 平成9年(1997)6月6日

(51) Int.CL*	施34配号	广内集理器号	F 1		技術表示個所
G 0 6 F 12/06	3 2 0	7623 ~ 5 B	G06F 1	2/08	320
		7623 - 5 B			J
		7623 - 5 H			н
3/06	302		3/06 3 0 2 A		
	304				304C
			等空網求	未請求	耐水項の数19 OL (全 14 頁)
(21)出票等等	特数平7~300967		(71)出額人	0000051	108
			t	株式会:	社自立類作所
(22) 出票 日	半成7年(1995)11月20日		1	東京都*	作代田区神田教育台門丁目 6 香地
			(72) 免明者	小井	MM
				神奈川	从小田原市区内体2550多数 株式会
			i	200	野作所ストレージシステム事業部内
			(72) 52 93 &	松本	OR-T-
			1	神奈川	从小田原市区府综2880条地 株式会
				HOR	個作所ストレージシステム事業部内
			(72) 発明者	村馬	2017
				神奈川	路小山原市西南岸2880春地 株式会
			1	₩ II 🗘	製作所ストレージシステム事業部内
			(74)代理人	弁理士	小川 獅勇

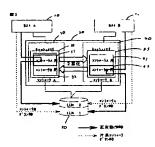
- 1 -

(54) 【発明の名典】 配管サブシステム

(57) [#49]

【課題】キャッシュを共有する複数のコントローラ間の キャシュの体色制御をなくした制御装置を有する記憶サ ノステムを提供する。

/// Aを投票する。。 【解決手段】互いに多重書きされているキャッシュ3 3、43において、キャッシュ領域を各プロセッサ毎に 分割し、各々のコントローラ30、40は、自コントロ - ラ餅醤エリアのみにアクセスする。 各コントローラが ・ ファット・ ファット・ ファット・ ファット・ ファット・ アー 使用する キャッシュ 領域 を固定化することにより、 プロセッサ間の球色制御を不要とし、 複数プロセッサ化に伴う性能劣化を防止することが可能となる。



2003 D4 24 10.53

【特許請求の範囲】

【雑字取1】 ホストコンピュータのデータを終終し、複 数の記憶衛城を有する記憶装置と、

故ホストコンピュータの指示に基づいて故紀像装置の納 御を行い、彼ホストコンピュータとはディスク装置との 関のデータ転送を制御し、故ホストコンピュータとは記 機装置との間を転送されるデータを一時的に保持する役 数の値域を有するキャッシュメモリを有する役数のコン トローラと前記複数のコントローラ間を接続するパスと **を具備する制御装置とそ有する記録サブシステムであっ**

前記コントローラには、前記記憶装置の複数の記憶領域 のうち少なくとも1つと数コントローラのキャッシュメ モリの複数の領域のうち少なくとも1つと終記パスによ り接続される他のコントローラのキャッシュメモリの程 数の領域のうち少なくとも一つが割当てられることを特 数とする記憶サブンステム。

【請求項2】 請求項1記録の記憶サブシステムにおいて、前記コントローラは、前記ホストコンピュータから 転送されるデータを放コントローラに割当てられている 複数の約記キャッシュメモリに書込むことを特徴とする 紀像サブシステム。

【継求項3】 継求項2記載の記憶サブシステムにおい で、前起コントローラに降客が発生したときは、前記也 のコントローラは鉄整客コントローラが担当していた値 紀記憶装置の記憶領域の処理を行うことを特徴とする紀 使サブシステム。

「独求項4」 独求項目を促の起処サブシステムにおい て、鉄色のコントローラはホットスタンパイしているコ ントローラであって、ホットスタンパイしているコント - うには、キャッシュメモリの記憶領域を割り当てな いことを特徴とする記憶サブシステム。

【請求項5】請求項1記数の記憶サブシステムにおいて、前記制御装置は推放の前記コントローラ間を接続す るパスをなし、自紀コントローラが他のコントローラに **新当てられた前記記像装置の記像領域に対する処理要求** をホストコンピュータから受取ったときは、前記コント ローラは、前記他のコントローラに放処理要求を通信することを特徴とする記憶サブシステム。

【請求項6】 請求項1記載の記憶サブシステムにおい て、前記キャッシュ領域の分割は、コントローラの負荷 に応じて変更することを特徴とする記憶サブシステム。 { 翻求項7} ホストコンピュータのデータを格納する複 数の論理ポリュームを有する磁気ディスクと、

誰ホストコンピュータと数ディスク装置との間を転送さ れるデータを一時的に保持する複数の領域を有するキャ ッシュメモリと、前記キャッシュメモリとが接続され、 技データのデータ転送を制御するデータ転送制御座とを 有する複数のコントローラと、複数のコントローラ間を 基づいて鉄磁気ディスク装置の制御を行う制御装置とを なする記憶サブシステムであって、

れずした版ファンステムというで、 前記コントローラには、前記邸気ディスク装置の複数の 森理ポリュームのうち少なくとも1つと鉄コントローラ

のキャッシュメモリの複数の領域のうち少なくともトつ と、他のコントローラのキャッシュメモリの複数の領域 のうち少なくとも!つとが割当てられることを特徴とす

○配管リフンステム。 【請求項8】 請求項7記載の記憶サブシステムにおい て、心足コントローラは、心足ホストコンピュータから

転送されるデータを数コントローラに割当てられている 数コントローラのキャッシュメモリの領域と、数コント ローラに割当てられている他のコントローラのキャッシ ュメモリの保城とに書込むことを特徴とする記録サブシ

て、仮記コントローラに抑劣が発生したときは、前記包 のコントローラは経験省コントローラが担当していた前 記論年ポリュームの処理を行うことを特徴とする記憶サ プシステム.

【雑求項 1 0】 結求項 9 記載の記憶サブシステムにおい て、前記他のコントローラはホットスタンパイしている コントローラであって、、ホットスタンパイしているコ ントローラには、キャッシュメモリの記憶領域を割り当 てないことを特徴とする記憶サブシステム。

【請求項11】 請求項7記載の記憶サブシステムにおい て、前記コントローラが他のコントローラに担当てられ た論理ポリュームに対する凱摩修収をホストコンピュー タから受けとったときは、森紀コントローラのデータ転 洗剤器感は、放他のコントローラに自記第一のパスを介 して処理要求を転送し、放処理要求を受難した該他のコ ントローラが抜為理ポリュームに対する処理を行い、処 理結果を、前記コントローラに転送することを特徴する

【韓求項121 韓求項7 紀載の記載サブシステムにおい て、前紀キャッシュ領域の分割は、コントローラの負荷 に応じて変更することを特徴とする記憶サブシステム。 【趙求項13】 趙求項7記載の記憶サブシステムにおいて、前記コントローラ間のパスは、2つの前記コントロ

ーラを接続する第一のパスと、、数 2 つのコントローラ の組を接続する第二のパスを含むことを特徴とする記憶 サブシステム. 「趙宝斯」4) 趙宝斯 (3 記載の記憶サブシステムにお

いて、前記制御装置にコントローラを増設するときは、 前記コントローラの2台単位に増設することを特徴とす る記憶サブシステム。

【緯求項15】ホストコンピュータのデータを格納し、 複数の記憶領域を有する記憶装置と、

波ホストコンピュータの指示に基づいて鉄起像装置の制 接続するパスとを有し、彼ホストコンピュータの指示に 5m 御を行い、故ホストコンピュータと終記像装置との間の

2003 04 24 10.53

記憶サブシステム

特別平9-146842

データ転送を制御し、数ホストコンピュータと放記敷装 置との間を転送されるデータを一時的に保持する複数の 価値を存するキャッシュメモリを有する複数のコントロ 備する制御装置とを育する記憶サブシステムであって。 点記コントローラには、前記記憶装置の複数の記憶領域 のうち少なくとも1つと終コントローラのキャッシュメ そりの複数の領域のうち少なくとも1つと前記パスによ り接続される他のコントローラのキャッシュメモリの複 酸の個域のうち少なくとも一つが割当てられ、

前記ホストコンピュータから転送されるデータは、終コ ントローラに割当てられている終コントローラのキャッ シュメモリの価値と、数コントローラに割当てられてい る他のコントローラのキャッシュメモリの領域に書込まれることを特徴とする記憶サブシステム。

【糖求項16】糖求項15記載の記憶サブシステムにおいて、前記コントローラに降省が発生したときは、前記 他のコントローラは政障者コントローラが担当していた 前記記憶装備の記憶領域の処理を行うことを特徴とする

【糖求項】7】 請求項 1 5 紀載の記憶サブシステムにお いて、故色のコントローラはホットスタンパイしている コントローラであって、ホットスタンパイしているコン トローラには、キャッシュメモリの記憶領域を割り当てないことを特徴とする記憶サブシステム。

「神像師181 神像師15 記録の記憶サブシステんにお いて、自記制御答案は複数の背記コントローラ間を技統 するパスを有し、粒配コントローラが他のコントローラ に割当てられた前記記憶装装の記憶領域に対する処理数 求をホストコンピュータから受取ったときは、前記コン トローラは、自紀仏のコントローラにお処理を求を通信 することを特徴とする記憶サブシステム。

【静葉項19】 離束項15記載の記憶サブシステムにお いて、肩起キャッシュ領域の分割は、コントローラの負 荷に応じて変更することを特徴とする記憶サブシステ

(発明の属する技術分野) 本発明は、上位装金からの特 報の入出力要求を制御する制御装置を有する記憶サブシ ステムに関し、特に、製御協調内のコントローラ及び# ャッシュメモリを冗長構成とする記憶サブシステムに関

[0002]

【従来の技術】コントローラ及びディスク多の記憶装置 に冗長性を終たせた記憶サブシステムとして、

[0003]特殊平4-215142に記載されている 記憶サプシステムは、現用系のディスク装置の記憶体器 50 ムでは、この禁止制御により、シングルコントローラの

を由果からアクセス可能な共用ディスク装置を介して予 備系のディスク装置に複写すること、あるいは、規用系 コントローラ際省時は、予備系のコントローラによっ て、現用系のディスク装置の記憶情報を抽出可能とする ことによって、コントローラ及びディスク装置障害時の 05

データ保全性の向上を計っている。

[発明が解決しようとする課題] 最近の市場動向とし て、記憶装置の高性能化、大容量化、低価格化の要求が 高まっており、RAIDの技術が重複されている。RA

10 I Dの技術を適用した記憶サプシステムにおいては、複 数のディスク装置をアレイ状に構成する。そして、データ書き込み時には、書き込みデータに加えて冗長データ を書き込みデータを格納したディスク装置とは異なるデ ィスク装置へ書き込む。アレイ構成内の任意のディスク

装置の故障に対しは、他のディスク装置のデータと前記 冗長データから跨客ディスク装置上のデータを修復可能 とすることによって、ディスク装置のデータの保全性の 向上を計っている。

[0005] しかし、RAIDの技術を適用した記憶サ プシステムは、データの保全性が向上する反面、上述し た冗長データ生成。「書き込みのために処理時間が増大するため、ホストからの1 ()処理と同期して冗長データ の生成、書き込みまでを行うと、ホストからのライト作 能が大幅に劣化する、従って、RAIDの技術を適用し Z5

シュが不可欠となっている。 【0006】ライトキャッシュとは、コントローラ内に 構成された。データを一時的に書き込むキャッシュであり、ホストからのライトを求では、このキャッシュに書き込みを行った時点で、ホストに終了程告を返す。そし

た記憶サプシステムのコントローラには、ライトキャッ

て、ホストの 1 〇処理とは非同期に冗長データの生成、ライトデータ及び冗長データのディスク装置への格 類を行うことによりまき込み処理込む性能低下を描く。 35 しかし、ライトキャッシュを用いると、データをキャッシュ上におき込んだ時点でホストに終了報告をするた

の、キャッシュ上にディスク装置表皮軟のホストデータ が存在する。従って、キャッシュに允良性がなければ、 キャッシュ検客時に、コーザデータロストとなる。した がって、特にデータの高信頼性が強く求められる記憶サ プシステムに用いる制御装置では、従来のコントロー

ラ、記憶装置の冗技構成に加え、一般的にキャッシュに も冗長性を持たせることが行われている。 【0007】 コントロー うを多数化した記憶サブシステ ムにおいて、単にキャッシュを多重化すると、キャッシュ上のデータを複数の制御装置から同時にアクセスする ことによるデータ整合性不断を防ぐためにキャッシュア クセス時に収扱の新御袋置からの基地制御が必要になる。そして、コントロークを多様化した起始サブシステ

記憶サブシステム

特開平9-146842

記憶サプシステムに比べて性能が低下する。

【0008】 本発明の目的は、コントローラの多葉化及びキャッシュの多重化に伴う、プロセッサ間のキャッシ ュの揉他制御を無くし、性能を募とすことなく信仰性を (000091

(無数を解決するための手段)上記の目的を達成するた め、本発明による記憶サブシステムは、各プロセッサ毎 に領費担当の論理ポリュームを詳値的に決める手段と、 あるプロセッサが受領したホストコンピュータからの要求が、担当外であった場合は、担当プロセッサに処理要 東を通信する手段と、上記遺信を受明したプロセッサ は、処理結果を要求元プロセッサに通信する手段と、各 プロセッサ毎に、ディレクトリ、「データセグメント等の キャッシュ網成要案を持つ手段と、上記構成要素の状盤 をプロセッサの負荷に応じてダイナミックに変更する手 段と、ホストコンピュータからのライトデータを複数の コントローラ上のキャッシュへ多重響をする手段と、コ ントローラ報告時には、報告コントローラ内プロセッサ の持つキャッシュ組成要素の制御権を正常系コントロー う内のプロセッサに切り替える手段と、コントローラ復 田時には、放射調査を提出プロセッサに見す手段と、デ ィスク装置への書き込み時にキャッシュメモリ際占が発 生した際は、多重者をしている他キャッシュからディス

ク装置に書き込みを行う手段とを有する。 【0010】上述した手段により、複数のプロセッサ間 で、キャッシュを排他制御することなく複数のコントロ ーラ Lのキャッシュへ多重要をを行うことができ、複数 プロセッサ化に伴う性能低下の発生を防ぎ、性能を痛と すことなく信頼性の向上を討ることができる。

【0011】また、上述の手段により、キャッシュメモリ際省時には、多重要さしている他キャッシュからディ スク装置への書き込みを行い、データロストを防止でき

【0012】さらに、七紀手段により、コントローラ牌 害時には、自動的に、正常系に切り奏えて処理統行が可 能であり、また、コントローラ提相時には、自動的に、 福田系に処理を戻すことが可能となり、システムの無停 止運用を実現できる。

[0013]

[発明の実施の影響] 関1は、本発明の最急間である 【0014】凶1において、10.11はホストコンピュータ、20はデュアルコントローラ構成をとる対策値 型、50はディスク装置であり、ディスク装置50は、 会理ポリューム0と高度ポリューム1の2つの急呼ポリ 一人に分割されている。

【0 0 1 5】 ホストAIOは、納御答案20内のコン ローラス30を介して、森卑ポリューム0の色質を行っ ており、ホストBilは、納御保置20内のコントロー 100間のデータ転送を制御す ラB40を介して、森間ボリューム1の処理を行ってい 50 0、6000より構成される。

【0.0.1.6】 ここで、コントローラA30には論理ポリ ューム Oが、コントローラ R 4 D には森理ポリューム トが処理相当森理ポリュームとして割当てられている。

又、コントローラ内のキャッシュの組織は、それぞれ、 コントローラA用キャッシュ31、41、コントローラ B用キャッシュ32、42に2分割されている。そして、コントローラA用キャッシュ31と41の間で2重 重きを行い、又、コントローラB用キャッシュ32と4

2 の間でも2 重要さを行う。 [0017] コントローラA 3 0 は、通常、コントロー ラA用キャッシュ31と41を用いて、1, O処理を行 い、同様に、コントローラB 4 0 は、コントローラB用 キャッシュ32と42を用いて、1,1O処理を行う。こ

のように、コントローラ毎に使用するキャッシュ領域を 個別に割り当てることにより、コントローラ間の排他制 御を無くし、コントローラ台数増加に伴う性能劣化を妨

【0018】また、コントローラお40粒合時には、コ ントローラB用キャッシュ32、42をコントローラA 30が使用することにより、ホストA10からコントロ ーラA30を介して、コントローラB40の処理科当で あった白理ポリューム1への処理を統行させることがで

【0019】以下、本発明によるマルチコントローラ構 成の制御装置の1実施例を図面を用いて説明する。 【0020】図2は、本発明をマルチコントローラ構成 の磁気ディスクアレイサブシステムに適用した場合の構

成例である. [0021] M2ESUT. 1000. 1100. 12 00、1300はデータ見埋を行う中央処理協議であるホストコンピュータ、2000は不ルチコントローラ構成をとりディスク装置の制御を行う制御協議、700

0、7 | 0 0 はホストコンピュータのデータを格納する ディスク装置である。ここで、新御装置2000は、ホ ストパスに直動したスロットに乗し込みホスト位体内に 組み込む場合もあるし、制御装置として独立した使体に 組み込む場合もあるし、ディスク装置を組み込んだ収4 として実現する場合もある。また、ディスク装置群7 U O O 及び7 | O U は、データディスクとパリティディス

りからなるパリティグループを含んでいる。 さらに、デ ィスク装置杯7000は、 論理ポリュームロと論呼ポリ ューム!とに、ディスク装置群7100は森煙ポリュー

ム2と森理ポリューム3とに分割されている。 【0022】 制御装置2000は、ホストコンピュータ 1000、1100とディスク装置7000間のデータ 転送を制御するコントローラ3000、4000以びホ ストコンピュータ1200、1300とディスク装置7

1 0 0 間のデータ転送を無償するコントローラシロロ

特闘平9-146842

ータ1000とのプロトコル制御を行うホスト(下制 御部3100、コントローラ全体を制御するマイクロブ ロセッサ (以下 (プロセッサ) という。) 3200. デ ータの転送を実行するデータ転送制御服は300、ホストコンピュータ1000とディスク装置7000デー タ転送時及びプロセッサ間通信時に用いられるキャッシ ュ3400、告ディスク装置7000とのプロトコル制 器を行うDRVI、下割機能3500より構成される。

ーラ3000と同一の構成である。 【0024】プロセッサ3204は、後述の手段によ り、あらかどめプロセッサ毎に禁他的に努り当てた刑当 無理ポリュームの処理を行う、このプロセッサ毎の利当 無理ポリュームの指定は、ホストコンピュータから結理 ポリューム毎の担当プロセッサ指定コマンドを受け取る ことにより、ダイナミックに収定可能である、このプロ セッサと担当論理ポリュームとの対応情報は、後述のキ ャッシュ上の共通メモリ戦略3410、4410に移時

コントローラ4000、5000、6000はコントロ

【0023】コントローラ3000は、ホストコンピュ

【0025】データ転送制御部3300はプロセッサ3 200からの指示により、ホストコンピュータ 1000 からのライトデータを指定キャッシュに多重書きする機 新を確まている。この実施例の構成では、キャッシュ3 400とキャッシュ4400の間で2歳書きを行い。 た、キャッシュ5400とキャッシュ6400の間でも 2 重書きを行う。以下、キャッシュ3 4 0 0 とキャッシュ4 4 0 0 の 2 面に 2 重書きする方式について 説明す

100261 ****** 3400 **** ** 4400 の内容について図3を用いて説明する。 ぬキャッシュ3 400とキャッシュ4400は内部構成が同一であるた キャッシュ3400を例に説明する。キャッシュ: 400は、プロセッサ間通信に用いる新御情報を格納し ている共通メモリ領域3410、プロセッサ3200月 低域3480、プロセッサ4200用低域3490より 構成される.

(0027) プロセッサ3200用機械3480は、ホ ストコンピュータとディスク装置間のデータ転送時、デ ータを1次的に格納するデータ格納エリア3482、デ ータ格前エリア3482を管理するデータ管理体験34 81より構成され、データ格両エリア3482に格納するライトデータと、このライトデータの管理情報は、4 ヤッシュ4400内のプロセッサ3200角領域448 0に2度券をを行う。同様に、プロセッサ4200用領域3490は、プロセッサ4200により、キャッシュ 4 4 D D内のプロセッサ 4 2 D D 用領域 4 4 9 D のライトデータとライトデータの管理情報が2 重響をされてい

430、多重要を情報3450、プロセッサ間コミュニ ケーションメモリ3460より構成され、これらの情報 は全て、データ転送が観想3300、4300によっ て、キャッシュ3400と4400に2項目をされてい

【0029】 83 (c) にプロセッサ南コミュニケーシ ョンメモリの構成を示す。プロセッサ間コミュニケーションメモリ3460は、プロセッサ3200、420

0、5200、6200 6時の過ぎ込み用メモリ34セ 1、3462、3463、3464より構成される。 3(d)にプロセッサ書き込み用メモリの構成を示す。 プロセッサ3200番を込み用メモリ3461は、日フ ロセッサ以外のプロセッサ4200、5200、620

りへの要求用エリア3471、3472、3473と行 プロセッサ以外のプロセッサ4200、5200、62 00からの要求に対する必番用エリア3474、347 5、3476より構成される。プロセッサ4200、5 200、6200mを込み用メモリ3462、346 3、3464の内部構成は、プロセッサ3200mを込

み用くそり3461と同一構成である。 【0030】キャッシュ5400とキャッシュ6400 との関も、共通メモリ領域を除いて、キャッシュ340 りとキャッシュ4400との間と阿様に2重化が行われている。共通メモリ蜘蛛は、キャッシュ3400、44 UDCで重要さされている情報を制御装置内の全プロセ サで共有するため、キャッシュ5400、6400に

【0031】本発明を実施する制御装置では、コントロ ・ラの地及はコントローラ2と単位で行い、材になった コントローラのキャッシュ間のみで2面書きを行うとと もに、ドライブ側のデータバスについても、それぞれの ディスク装置は対になったコントローラにのみ技術する ことによりハードウェア構成を商略化し、ドライブ側デ ータバス上の競合を回避することが可能となる。

【0032】次に本実施例における、磁気ディスクサブ シスチムでの、ホストコンピュータ1000からの1. O処理について図4、図5、図6を用いて放明する。ま プロセッサ3200担当論様ポリュームへの 1,10処理について説明する。

【0033】図4は、ホストからの1、O処理を示すフローチャートである。ホストコンピュータ!000から の書き込み要求時、プロセッサ3200は、まず、共通 メモリ競戦3410内の論理ポリューム担当プロセッサ 情報3420によって、処理要求論理ポリュームの担当 プロセッサ情報を取得し、自処理担当幕理ポリューム (LUN)への処理かの利定を行い(ステップ90

2)、自プロセッサ処理担当論理ポリュームへの処理で る。 あることを建成する。次に、処理程例の何定を行い(ス 【0028】共通メモリ領域3410は、論理ポリュー 50 テップ903)、書を込み処理であることを提出する。

2003 04 24 10 53

ホスト1、F朝御部3100により、書き込み益均デー タを受難し、データ転送制御部3300によってキャッシュ3400のコントローラ3000別面破3480と キャッシュ4400のコントローラ3000用蜘蛛44 とにその管理情報とともに2歳に格納する(ステッ ブ904)、そして、この時点でホストコンピュータ1 000に終了を報告する(ステップ905)。 【0034】図5は、キャッシュ内のデータをディスク 装御に格明する処理を示すフローチャートである。プロ

セッサ3200は、ホストコンピュータ1000からの *O処理とは非同期にプロセッサ3200用額減34 50上のウイトデータをデータ転送制御配3300とD RV 1. F制理部3500によりディスク装置数70 00に格納する(ステップ922)。この数、キャッン 1のメモリ健実により並み込みエラーが発生した場合 (ステップ923) は、2重化しているプロセッサ32 00用銀載4480からディスク装置7000へ格納す る (ステップ924) ことによりデータ損失を妨止する ことができる。

【0035】ポストコンピュータ1000からの数み込 み要求時は、プロセッサ3200は、上記書を込み処理 **阿條、自プロセッサ処理担当会理ポリューム(LUN)** への処理であることを認識(ステップ902)した権。 処理種別の料定を行う(ステップ9 0 3)、1.1〇処理 が挟み込み処理であることを記載すると、データ転送制 構成3300とDRV 1、下制機成3500に上りデ ・タモディスク装置群7000からキャッシュ3 のコントローラ3000用類域3480に格納し 1ステ ップ906)、ホストコンピュータに転送する (ステッ

プリロ7)。 【0036】 本にホストコンピュータ1000からコン トローラ4000担当倫理ポリュームへの (、〇気度に

【0037】ホストコンピュータ | 000からの書き込 み要求時、プロセッサ3200は、まず、共通メモリ伽 城3410内の島煙ポリューム担当プロセッサ情報34 20によって、処理要求論理ポリュームの担当プロセッ サ情報を取得し、自幼専組当業理がリュームへの処理かの料定を行い(ステップリロ2)、処理和当外論理がリ コームへの処理であることを試験する。次に、処理種別 の料定を行い(ステップ908)、書き込み処理である ことを認識する。そして、ホストコンピュータ1000 からの書き込み論理データをキャッシュメモリのコント ローラ3000用領域3480に格納し、書き込み処理 そこの論理ポリュームの担当であるコントローラ400

0へ要求する (ステップ 9 0 9) 。 (0 0 3 8) プロセッサ3 2 0 0 は、プロセッサ4 2 0

3 3 0 0 により共通メモリ領域 3 4 1 0 、 4 4 1 0 内の プロセッサ32008を込み用メモリ内のプロセッサ4 200への要求用エリアに2項に格納する。ここで、数 理権別情報とは、最近込み処理が扱み込み処理がを判断 する情報である。プロセッサイ200は、例えば10m s といった ・定時間で、共通メモリ蜘蛛3.4.1.D.

1 Uの自プロセッサへの复采用エリアを基別にいき、他 プロセッサからの要求を認識する。 [0039] 内6は、プロセッサ3290からの処理者

本を受視したときのプロセッサイ2200からの処理を ボを受視したときのプロセッサイ200の処理をポすプ ローチャートである。和述の刀瓜により、プロセッサイ 200からの要求を提賞(ステップタよ1)したプロセ ッサイ200は、プロセッサ3200歳を込み用メモリ 内のプロセッサイ200への要求用エリア内の処理権別

を参照し、春き込み処理会求であることを必定する (ステップ932)、そして、プロセッサ4200は、プロセッサ420では、プロセッサ420でロセッサ420 0への要求用エリア内の書き込み益理アドレス。書き込みデータのキャッシュ上の格納アドレス、データ長を取 得し (ステップ933)、キャッシュ3400内の数格

納アドレスからデータ長分の書き込みデータをプロセッ サ4200用領域3490と4490に、その管理情報 である書き込み論理アドレスとデータ員と共に、2重に 格明する(ステップ934)。そして、終了情報を共適 メモリ領域3410、4410内のプロセッサ4200 直を込み用メモリ内のプロセッサ3200からの要求に 対する応答用エリアに設定することにより、プロセッサ 3200に処理終了を通信する(ステップリ35)。

【0040】プロセッサ3200は、プロセッサ420 のに対する処理要求後は、プロセッサ4200番を込み 州メモリ内のプロセッサ3200からの要求に対する店 香用エリアを参数することにより、プロセッサイ200 の処理の終了を監視 (ステップ9:0) しており (図4

の鬼味の終了を監視(ステップリ 1 U) しており パロマ 参照)。この処理終了の適信を受けて、ホストコンピュ 一多!000に終了を規密する(ステップ905)。プ ロセッサ4200は、この後、図5に従ってホスト!。 〇処理とは非国際に、この書き込みデータのディスク協 置7000への番き込み処理を行う。 【0041】例4において、ホストコンピュータ100

0から読み込み要求があったときは、プロセッサ3200は、上記書き込み要求受額時向様、処理担当外籍理求 リューム (LUN) への処理であることを認識した (ス テップ902) 後、処理権別の料定を行う (ステップ9 08) 。 読み込み処理であることを認識すると、プロセ ッサ3200は扱み込み要求必定アドレス、扱み込みデ ータのキャッシュ上の格納許可アドレス、データ長、処 理権別情報を共進メモジ領域3410、4410内のフ 0に著名込み処理を要求するために、、最名込みデータ ロセッサ3200番名込み用メモリ内のプロセッサン は使了ドレス、着名込みデータのキャッシュ上の格納フ 00への音楽用エリアに格納することにより、数1.UN ドレス、データ長及び処理権別情報をデータ転送制書館 50 処理担当であるプロセッサ4200に放み込み要求を過

配信サブシステム

特別平9-146842

信する (ステップ911)。

【0042】図6において、プロセッサ3200からの 要求を認識(ステップ931)したプロセッサ4200 は、共通メモリ組織内の情報により、扱み込み処理であ ることを認識する (ステップ932)。 そして、共通メ モリ領域から読み込み要求論理アドレス、読み込みデー タのキャッシュ上の機動的リアドレス、データ技を収荷 する(ステップ936)、次に、データモディスク装置 7000からプロセッサイ200附領域イイ90に移向 し、このデータもキャッシュ3400上の格別お可アドレスに格納する(ステップ937)。さらに、共通メモリ領域3410、4410内のプロセッサ4200番を 込み用メモリ内のプロセッサ3200からの要求に対す る応答用エリアに終了情報を設定することにより、プロ セッサ3200に鉄み込み終了を通信する(ステップ9

【0043】例4において、プロセッサ4200の処理 美了を監視 (ステップ912) していたプロセッサ32 00は、この親み込み終了報告を受けて、データをホストコンピュータに転送する(ステップ913)。

【0044】 このように、プロセッサ3200は、通 常、プロセッサ3200用類域348Dと4480を用いて、1. (O処理を行う、同様に、プロセッサ4200 は、通常、プロセッサイ200用的雇3490と445 0を用いて、1、10処理を行う。 {0045}このように、プロセッサ毎に使用するキャ

ッシュ領域を固定化することにより、プロセッサ間の鉄 数制御を無くし、プロセッサ台数増加に伴う性能劣化を 妨ぐことができる。特にホストコンピュータ間でファイ ル (急度ポリューム) モシェアしないシステムにおいて は、接続しているコントローラ内のプロセッサにこの会 様ポリュームを割り当てておくことにより、1.10何坪 のときのプロセッサ間の通信制御を不安とし、さらなる 性能向上を可能とする。

(0046) 次にコントローラ4000の設置時の自動 切り巻大、復形方式について図7、図8を用いて返明する。1、〇気確実行中、コントローラ4000の向方を 検知したプロセッサは、共通メモリ飢壊3.4.1.0を用い て、残りの全プロセッサにコントローライ000の応告 を適切する。この際、コントローライ000とキャッシ ュモ2重要をしているコントローラ3000内のブロヤ ッサ3200には、処理の引命要求も適保する。 4次編 例では、プロセッサ3200か配害を検知した場合につ いて収明する。

(0047) 例7は、プロセッサ3240がコントロー 4000の脚省を検知した場合のプロセッサ3200 の航費を示すフローテャートである。プロセッサコ20 0は1、「○気理実行中(ステップ950)、コントロー ラ4000の解析を使知(ステップ951)すると、森 迷の方法により、プロセッサうでもり、6200にコン

トローラ4000の障害を通信する。そして、障害コン トローラモシステムから切り放すため、キャッシュ34 00と4400へ2重奏きされているホストコンピュー タからの書き込みデータ及び共通メモリ領域のデータ を、キャッシュ3400への1重素さに変更することを データ転送制御練3300に指示する (ステップ95 2)。また、プロセッサ3200からの世末を認識した プロセッサ 5 2 0 0 、 6 2 0 0 は、共通メモリ領域をキャッシュ 3 4 0 0 への 1 仮書きに変更する。次に、プロ セッサ3200は、プロセッサ4200の処理を引き数 ぐ為に、プロセッサ4200用額域の納御権をプロセッ サ3200に切り着える(ステップ953)。これらの

処理により、制御権の切り替えが完了し、プロセッサ3 200は通常の1、〇処理を再開する(ステップ95 [0048] 図8は、数字が発生したコントローラ40

0 0の復旧処理を示すプローチャートである。コントローラ4000の際害庭位が交換(ステップ971)され ると、プロセッサ 4 2 0 0 は、共通メモリ領域 3 4 1 0 20 を用いて全プロセッサに復旧開動を伝達する (ステップ 972) . JDt > # 3200, 5200, 6200

この復旧開始の伝達を受けて (ステップ955) それぞれのコントローラのデータ転送網御館にキャッシ ュ3400と4400への2歳書きを指示すると共に、 共過メモリ領域3410、4410を用いて、処理終了

の応答をプロセッサ4200に通信する(ステップ95 6)、この終了場合を全プロセッサから受難(ステップ 973)したプロセッサイ200は、キャッシュ440 0のデータ同復を行う(ステップ974)。データ回憶 が完了すると、共雄メモリ領域3410、4410を用いて、プロセッサ3200に提出完了を伝達する(ステ

v7975). 【0049】関7において、この完了適知を受けた(ス トップ9ち8)プロセッサ3200は、プロセッサ42

0 り削削減の制御後をプロセッサイ2 0 0 に似田(ステップ9 5 9)させ、共通メモリ削減を用いて、制御格の 収制をプロセッサイ2 0 0 に伝達する (ステップ9 6 り) 、M8において、この伝達を受けた(ステップリア プロセッサ4200は、1 '〇粒理を削削させる (ステップタ77)。

【0050】 森、以上の実施側においては、コントロー ラ毎にプロセッサ、ホストリ、F制製器を1つ作った例 を示したが、これらの数は任意でも、ホストコンピュー タからのコマンドを受け取ったプロセッサが、利当プロ セッサに処理要求を促進することにより、同様に実現で

100511 ft fembrastate Tee が毎に均等ではなく、ユーザの指定により設定 変更可 成である。特に、特定ツロセッサをホットスタンパイで 50 動作させる場合には、キャッミュ和城をホットスタンパ

2001 04 24 10 51

記憶サブシステム

特別平9~146842

イのプロセッサには削り当てないことにより、キャッシ ュモ有効に利用することができる。 X. プロセッサの負 毎に応じてダイナミックに変更することも可能である。 ユーザの指定により分割を行うか、プロセッサの負荷に 応じて変更を行うかの指示は、本実施例では、ホストコ マンドにより行うが、パネルといった装置を接続し、そ

こから入力する形を取っても、むろん良い。 【0052】 つぎに、コントローラのキャッシュの動的 割当の実現方式について、以下、説明する。まず、キャ ッシュの管理方式について、図9を用いて規則する。 【0053】プロセッサ毎に持つデータ格納エリアは、 セグメント983と呼ばれる管理単位に分割されてい る。セグメントは、セグメント毎にセグメント管理プロック981(以下SGCBという。)をデータ管理情報 った。 内に持ち、セグメントを管理する情報とセグメントアド レスが格朝されている。又、これらのSGCRは、その セグメントの属性によって、ダーティキュー980とク リーンキュー9 8 2 という 2 つキューに分けられて接続 されている、ダーティキュー9 8 0 には、ディスク未反 乗のライトデータを修飾しているセグメントのSGCB

【0054】キャッシュの動的製造を実現するために、 プロセッサ毎の負荷情報を共通メモリ領域に持つ。この 負荷情報として、例えば、キャッシュ内のクリーンSG CB量を用いる。各プロセッサは、SGCBのクリー ン、ダーティ間のキュー遷移契機に、この情報を更新する。プロセッサは、何太ば、1分といった一定周期でこ の情報を参照にいき、キャッシュを共有しているプロセッサ内で最も負荷の低いプロセッサ内のリーンキューか ら最も負荷の高いプロセッサのクリーンキューへ、その 色数が同じになるまでSGCBと養辣セグェントをは行 させる。この際、使用中のSGCBは、移行対象外とす る。移行の際は、SGCBの格納データ情報はクリアす る。この移行の間は、プロセッサ通信を用いて、移行を 行うプロセッサの 1、「O 処理はとめる。

が接続されており、それ以外のSGCBは、クリーンキ

-982に接続されている。

【0055】また、以上の実施例においては、2分のコ ントローラ間でキャッシュを共有し、各々、 科コントロ ーラのキャッシュに 2 重書きする例を示したが、キャッ シュ価値がプロセッサ毎に分裂されていれば、そのキャ ッシュの共有化方式。多数者さが式は、任意の方式でも、同様に実現できる。

【0056】キャッシュ多数書きの例を図10に示す。 (1) は、装置全体でキャッシュを共作しあい、2重値 きずる方式である。つまり、プロセッサ3200はキャ マシュ3400、4400を用いて、プロセッサ420 0はキャッシュ4400、5400を用いて、プロセッサ420 75200はキャッシュ5400、6400を用いて、 プロセッサ6200はキャッシュ6400、3400を 用いて2素者をを行っている。

【0057】 (2) は、協議会体でキャッシュを共有し あい、全キャッシュに多重書きする方式である。つま り、プロセッサ3200、4200、5200、620 Dは、それぞれキャッシュ3400、4400、540 の、6400を用いて、多数書きを行っている。このケ 一スにおいて、コントローラが障害となった場合は、キ ヤッシュを共有しているプロセッサ間でもっとも負荷の 低いプロセッサが、稼ぎコントローラ担当論理ポリュー

ムの処理を引き継ぐ。これらのケースにおいては、任意 のプロセッサが障害コントローラ担当森町ポリュームの 処理を引き続けるように、ディスク側のデータバスを、 装置内の全ディスク装置、全コントローラで共通のバス に技統しておく。もちろん、これらの多重書き方式を装 関内で設在させることも可能である。これらの多重書き

方式の指定は、共通メモリ酸域3 4 1 0 、 4 4 1 0 に多 最審さ情報を持ち、各々のプロセッサ3 2 0 0 、 4 2 0 0. 5200、6200が、この情報を元に、進き込み データの転送方式をデータ転送割提課3300 0、5300、6300に指示することにより実現でき 20 გ.

【兒明の効果】本兒明によれば、コントローラ及びキャ ッシュメモリを2番化した記憶サブシステムにおいて、 各コントローラにキャッシュメモリの「部及び毎晩ポリ ュームを割り当てることによりキャッシュメモリに対す るコントローラ内のプロセッサ間の建約制御が無くなる ため、複数プロセッサ化による応答性能劣化を防ぐこと

【0059】また、複数のキャッシュへ多重響きするこ とにより、キャッシュ教育時には、多重書きしている他 キャッシュからディスクに書き込むことができるため、 データロストを防ぐことができる。さらに、コントロー う数週時にキャッシュメモリの制御を正常なコントロー うに切り替える手段とコントローラ際資から提出する毛 段を設けることにより、システムを無停止で適用するこ

とができる. INGO DO UN TO DO

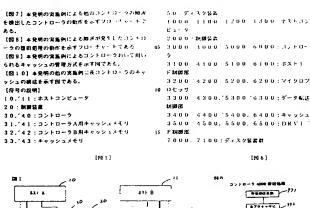
(凶!) 4兒明の概要を表す講成以である。 【似2】 4 発明の実施例である新藤装置の構成料であ

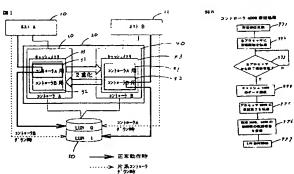
【阿3】 本発明の実施例であるコントローラのキャッシ の構成を示す物である。

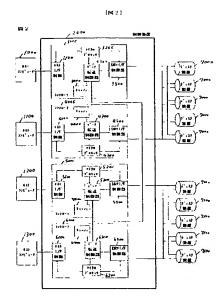
【関4】 4発明の実施例によるコントローラのホストか

らの1 O質理の動作を示すフローチャートである。 【図5】 本発明の実施例によるコントローラのキャッシ っ内のデータモディスク装置に集論する動作を示す ーチャートである。

【残ら】 4発明の実施例による色のコントローラからめ 理長求を受けとったコントローラの制御装置の動作を示 50 すフローチャートである。







2003 04 24 10 53

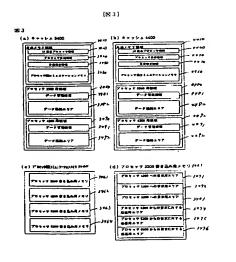
2003 04 24 10 53

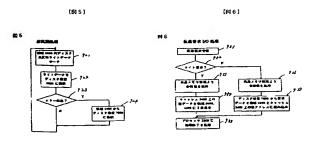
記憶サプシステム

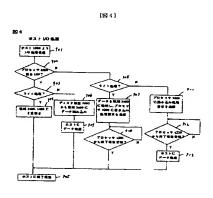
特別平9-146842

記憶サブシステム

特開平9-146842

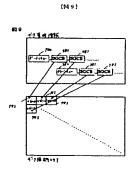






- 12 -

- 10 -

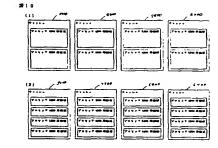


(図7)

| 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100

- 11 -

(MIO)



2003 04 24 10.53

- 14 -

2003 04 24 10.33